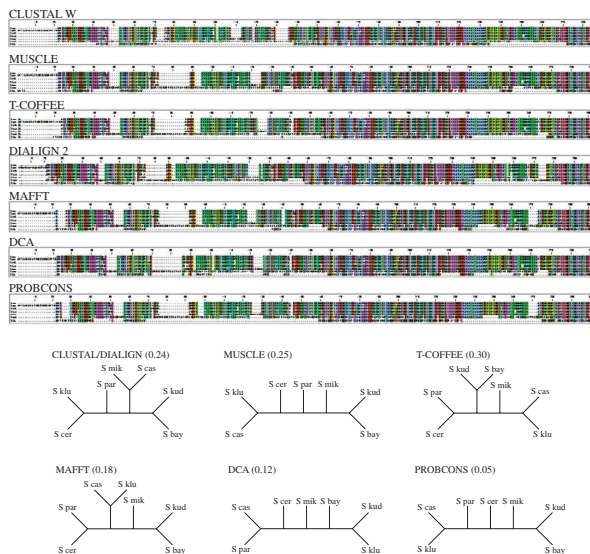


10. Filogenetika

AZ EVOLÚCIÓS TÖRZSFA az öröklések történetét mutatja. A fa gyökerénél a közös ős van, a leveleknél az élő leszármazottak, a belső pontok pedig köztes közös ősöknek felelnek meg. Az élekhez szokásosan valamilyen hossz tartozik, ami klasszikusan az időnek felel meg. Molekuláris evolúciós módszerek homológ szekvenciák hasonlósága alapján építik meg a törzsfát, ami ekkor a szekvenciák történetét mutatja. Hogy milyen homológ szekvenciákat válasszunk ki ahhoz, hogy az egyben az organizmusok történetének is megfeleljen, az egy külön tudomány. Például szokás riboszomális RNS alapján építeni a fajok törzsfáját, illetve „alapvető” fehérje-kódoló gének alapján. A ma élő összes élőlények genetikai diverzitása viszont igen nagy, és vitatható, hogy vannak-e egyáltalán olyan gének amik hűen tükrözik az egyedfejlődést¹, illetve beszélhetünk-e egyáltalán az Élet Univerzális Fájáról, tekintve, hogy laterális gén transzfer elég gyakori².

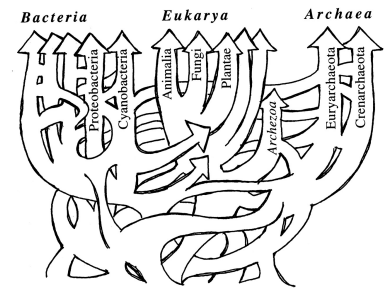
Törzsfák építése

Törzsfá építése valamilyen szekvenciák alapján történik, leggyakrabban első lépésként azokat illesztve. Megjegyezzük, hogy modern módszerek gyakran ennél is tovább mennek, mivel maga az illesztés egyáltalán nem egyértelmű³.



Ezen figyelmeztetések után elkezdjük a módszerek tárgyalását.

- ★ irányított (gyökeres) és nem irányított fa
- ★ lehetséges törzsfák száma
- ★ törzsfá építési módszerek: távolságok alapján (*distance-based*), parszimónia, valószínűségi módszerek (maximum likelihood).
- ★ Newick formátum



1. ábra. Doolittle (1999) illusztrációja, hogy hogyan érdemes elképzelni az univerzális fát.

¹ Tal Dagan and William Martin. The tree of one percent. *Genome Biology*, 7:118, 2006

² W. Ford Doolittle. Phylogenetic classification and the universal tree. *Science*, 284:2124–2128, 1999. DOI: 10.1126/science.284.5423.2124

³ Karen M. Wong, Marc A. Suchard, and John P. Huelsenbeck. Alignment uncertainty and genomic analysis. *Science*, 319:473–476, 2008

2. ábra. Wong (2008) példája, hogy 7 program 7 különböző illesztést ad ki ugyanazon homológ fehérjeszekvenciákra, ami 6 különböző törzsfának felel meg. Nincsen biztosan helyes illesztés!

Gyakorlat (9 pont)

Eheti gyakorlat: törzsfá készítése.

A gyakorlat célja, hogy valamilyen ortológ géncsalád alapján megállapítsuk a baktériumtörzseink (Chry5 és AT-6) evolúciós rokonságát többi referencia törzssel. A genomokat és a családokat az alábbi táblázatok mutatják:

Genomok		Géncsaládok	
Rövidítés	Teljes név	Kód	K84 azonosító
Chry5	<i>Agrobacterium tumefaciens</i> Chry5	recQ	ACM27148.1
AT6	<i>Agrobacterium vitis</i> AT6	rpoB	ACM26272.1
Ach5	<i>Agrobacterium tumefaciens</i> Ach5	rpoC	ACM26273.1
F2	<i>Agrobacterium tumefaciens</i> F2	SMc00019	ACM25745.1
H13_3	<i>Agrobacterium tumefaciens</i> H13-3	aroB	ACM27982.1
K84	<i>Agrobacterium radiobacter</i> K84	atpD	ACM28041.1
S33	<i>Agrobacterium tumefaciens</i> S33	dnaK	ACM24959.1
S4	<i>Agrobacterium vitis</i> S4	glyS	ACM25671.1
Bsui	<i>Brucella suis</i> 1330	leuS	ACM28526.1
Mrad	<i>Methylobacterium radiotolerans</i> JCM2831	lysC	ACM27909.1
Retl	<i>Rhizobium elii</i> CFN 42	murC	ACM27047.1
Rleg	<i>Rhizobium leguminosarum</i> bv. <i>trifolii</i> WSM1689	prfC	ACM25219.1
Sme1	<i>Sinorhizobium meliloti</i> 1021	secA	ACM27922.1
		thrA	ACM26416.1
		truA	ACM25314.1

- **a. (4 pont)** A listán a K84 genom néhány housekeeping génje szerepel: válassz ki egyet és keresd ki az ortológjait a másik 12 genomban. A 13 fehérjeszekvenciát gyűjtsd össze egy FASTA fájlban. A fájl nevében használd a család kódját a táblázatból.

☞ Töltsd le a kiválasztott géned fehérjeszekvenciáját a Genbankból.
 ☞ Az ortológokat megtaláld blastp kereséssel (a legkisebb E-value tartozik hozzájuk). Ezt alapvetően két módon lehet megcsinálni:
 (1) parancsból végrehajtod a blastp-t (~genome/bin könyvtárban megtaláld) az adott génekre a Chry5-ben és AT6-ban megkeresve a homológokat, illetve az NCBI blast interfészen keresztül a másik 10-et — jelöld be keresett organizmust a megfelelő mezőben (kalandos kedvűek letölthetik a prote.in.faa.gz fájlokat a teljes genomokra, megépíthetik a blastp adatbázist (makeblastdb) és helyben futtathatják a blastp-t, de ezt ne a bigyon csináld inkább, mert lehet, hogy túlságosan leterhelheti), vagy
 (2) a bigyon megtaláld az összes AT6/Chry5 blastp keresés eredményét (~genome/data/ könyvtár alatt), ahonnan ki tudod szedni az azonosítókat, ami alapján direktben kigyűjtheted a szekvenciákat a Genbankból.

- **b. (2 pont)** Illeszd a szekvenciáidat és írd ki az eredményt egy második FASTA fájlban. Erre használd a muscle programot (muscle -in nemillesztett.fa -out illesztett.fa), amit megtalálsz a ~genome/bin könyvtárban. A szekvenciák nevét írd át kézzel rövidebbre a Fasta fájlban, a fenti táblázatban megadott genom rövidítések szerint.
- **c. (2 pont)** Építsd meg a génszekvenciák szerinti törzsfát a PhyML programot használva, Dokumentáld, hogy milyen paramétereket választottál (figyelj a szekvencia típusra és a szubsztitúciós modellre). Ehhez a muscle outputját át kell konvertálnod a „Phylip interleaved” formátumba. (Megtetheted a http://www.hiv.lanl.gov/content/sequence/FORMAT_CONVERSION/form.html oldalon, kézzel, vagy ld. a bónusz feladatot.) A PhyML programot futtathatod online a <http://www.trex.uqam.ca> oldalról, vagy a bigyon a parancssorban.
- **d. (1 pont)** Rajzold ki a megépített törzsfát a Phylodendron programmal (<http://iubio.bio.indiana.edu/treeapp/treeprint-form.html>), és mentsd el az eredményt PDF formátumban.
- **c. (+2 bónusz pont)** Írjál egy Perl programot a FASTA → Phylip konverzióra. A Phylip formátumban az első sor 2 értéket tartalmaz: az illesztett szekvenciák számát (n) és az illesztés hosszát, szóközzel (vagy –közökkel) elválasztva. A következő n sor megadja a szekvenciák nevét és az illesztés első blokkját. A nevet kötelezően 10 karakter hosszan kell írni (ha hosszabb akkor szóközöket hozzá adni — substr(name," ",0,10)),

utána jöhet egy TAB, majd a szekvencia. A szekvencián belül tetszőlegesen lehet szökőzőket írni, hogy jobban nézzen ki (szokás minden 10. után), de nem fontos. Az első blokkot követheti akármennyi további blokk, amíg nincs vége az illesztésnek, de ott már nem kell a szekvenciák nevét újra megadni. Blokkok közé lehet üres sorokat írni. Részletesebb leírás: <http://evolution.genetics.washington.edu/phylic/doc/main.html#inputfiles>. (A legegyszerűbb megoldás az, ha csak egy blokkba írod a szekvenciákat, azaz először beolvasod a FASTA fájlt teljesen, aztán kiírod 1-1 sorba az illesztett szekvenciákat.)

```

6 39
Archaeopt CGATGCTTAC CGCCGATGCT
Hesperorn CGTTACTCGT TGTCGTTACT
Baluchith TAATGTAAAT TGTTAATGTT
B. virgin TAATGTTCGT TGTTAATGTT
Brontosau CAAAACCCAT CATCAAACC
B.subtili GGCAGCCAAT CACGGCAGCC

TACCGCCGAT GCTTACCGC
CGTTGTCGTT ACTCGTTGT
AATTGTAAAT GTTAATTGT
CGTTGTAAAT GTTCGTTGT
CATCATCAA ACCCATCAT
AATCACGGCA GCCAATCAC

```

Beadandó: páronként/csoportonként egy (1) dokumentum az előadáshoz kapcsolódó saját jegyzetekkel (1 oldal, 1 pont), a FASTA fájlok az **a** és **b** pontokból, a PhyML outputja (a Newick formátumú fa) a **c** pontból és a rajz (PDF) a **d** pontból. Ezt íméjl csatolmányban kell elküldeni nekem, legkésőbb 2016. április 25-én 14:59-ig.